DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 15 | Issue 1 Article 36

5-1-2016

Liu-Type Logistic Estimators with Optimal Shrinkage Parameter

Yasin Asar
Necmettin Erbakan University, yasar@konya.edu.tr

Recommended Citation
Asar, Yasin (2016) "Liu-Type Logistic Estimators with Optimal Shrinkage Parameter," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 1 , Article 36.
DOI: 10.22237/jmasm/1462077300
Available at: http://digitalcommons.wayne.edu/jmasm/vol15/iss1/36

Journal of Modern Applied Statistical Methods
May 2016, Vol. 15, No. 1, 738-751.

ISSN 1538 − 9472

Liu-Type Logistic Estimators with Optimal Shrinkage Parameter

Yasin Asar
Necmettin Erbakan University
Konya, Turkey

Multicollinearity in logistic regression affects the variance of the maximum likelihood estimator negatively. In this study, Liu-type estimators are used to reduce the variance and overcome the multicollinearity by applying some existing ridge regression estimators to the case of logistic regression model. A Monte Carlo simulation is given to evaluate the performances of these estimators when the optimal shrinkage parameter is used in the Liu-type estimators, along with an application of real case data.

Keywords: Logistic regression, multicollinearity, maximum likelihood, MSE, Liu-type estimator

Introduction

It is a very common problem to deal with highly intercorrelated explanatory variables. Especially in economics and in other applied research areas, the variables used in the multiple linear regression models are collinear. This problem is called multicollinearity. There are some results of the multicollinearity problem such as having inflated variance scores and instable estimations of the parameters when the ordinary least square (OLS) estimator is used. Similarly, in the logistic regression model, if the maximum likelihood method is used, these drawbacks occur at all. Also, one cannot obtain decisive answers to the related questions in both of the models.

There are some methods to deal with this problem. One method is to use ridge regression, first introduced by Hoerl and Kennard (1970). The other methods are to use Liu or Liu-type estimators proposed by Liu (1993) and Liu (2003) respectively.

These methods have been applied in the case of multiple linear models. However, there is not much attention paid to the multicollinearity problem in the case of logistic regression. Månsson and Shukur (2011), Kibria, Månsson, and

Mr. Asar is an Assistant Professor in the Department of Mathematics-Computer Science. Email him at: yasar@konya.edu.tr or at yasinasar@hotmail.com.

Shukur (2012), and Inan and Erdogan (2013) are some exceptions. In the first two studies, the authors used some early defined ridge estimators in the logistic regression model. In the last one, the authors applied Liu-type estimators given in Liu (1993) to the logistic model as well.

There is another Liu-type estimator defined by Huang (2012) for the logistic regression model. The author explained the theoretical advantages of this estimator and gave some comparisons. In this study, we use this estimator with optimal shrinkage parameter and some existing ridge parameters in order to make a simulation study to see the performance of these estimators in the logistic regression model.

Methodology

Consider the binary logistic regression model, a widely used method in statistical analysis, such that the dependent variable is Be($\mathbf{P}$) where $\mathbf{P} = e^{\mathbf{X}\boldsymbol{\beta}} / 1 + e^{\mathbf{X}\boldsymbol{\beta}}$ such that $\mathbf{X}$ is the design matrix of order $n \times (p + 1)$, p is the number of explanatory variables, and $\boldsymbol{\beta}$ is the coefficient vector of order $(p + 1) \times 1$. The most commonly applied method of estimating $\boldsymbol{\beta}$ is the maximum likelihood estimation (MLE) method.

One can compute the MLE of $\boldsymbol{\beta}$ by using the iteratively re-weighted least square algorithm as follows:

$$\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{W}\hat{\mathbf{z}} \ , \tag{1}$$

where $\mathbf{W} = \mathrm{diag}\left(P_i\left(1-P_i\right)\right)$ and $\hat{z}_i = \log\left(P_i\right) + \dfrac{y_i - P_i}{P_i\left(1-P_i\right)}$ is the i^{th} element of the vector $\hat{\mathbf{z}}$, $i = 1, 2, \ldots, n$.

The mean square error (MSE) of the MLE is given as follows:

$$\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right) = \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} - \boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} - \boldsymbol{\beta}\right)\right] = \mathrm{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} = \sum_{j=1}^{p+1} \frac{1}{\lambda_j} \ , \tag{2}$$

where λ_j is the j^{th} eigenvalue of the matrix $\mathbf{X'WX}$.

If some of the eigenvalues are small (close to zero), then the asymptotic variance of MLE becomes inflated. In other words, multicollinearity between the explanatory variable makes this estimator instable. Schaefer, Roi, and Wolfe (1984) proposed the following logistic ridge estimators to cure this problem:

$$\hat{\boldsymbol{\beta}}_k = \left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} \ , \tag{3}$$

where

$$k = \frac{1}{\hat{\boldsymbol{\beta}}'_{\mathrm{MLE}}\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}}, \frac{p}{\hat{\boldsymbol{\beta}}'_{\mathrm{MLE}}\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}}, \frac{p+1}{\hat{\boldsymbol{\beta}}'_{\mathrm{MLE}}\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}}$$

The following logistic Liu estimator $\hat{\boldsymbol{\beta}}_d$ was defined in Månsson, Kibria, and Shukur (2012):

$$\hat{\boldsymbol{\beta}}_d = \left(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{I}\right)^{-1} \left(\mathbf{X}'\mathbf{W}\mathbf{X} + d\mathbf{I}\right)\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} \ , \tag{4}$$

where $0 < d < 1$. Also, Huang (2012) defined the estimator $\hat{\boldsymbol{\beta}}(k,d)$ as a combination of the two different estimators given above such that

$$\hat{\boldsymbol{\beta}}(k,d) = \left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1} \left(\mathbf{X}'\mathbf{W}\mathbf{X} + kd\mathbf{I}\right)\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} \ , \tag{5}$$

where $k > 0$, $0 < d < 1$. It was shown that: if $d = 1$, then $\hat{\boldsymbol{\beta}}(k,d) = \hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$; if $k = 0$, then $\hat{\boldsymbol{\beta}}(k,d) = \hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$; and if $k = 1$, $\hat{\boldsymbol{\beta}}(k,d) = \hat{\boldsymbol{\beta}}_d$. In Månsson et al. (2012), it was proved that, when $d < 1$, $\left\|\hat{\boldsymbol{\beta}}_d\right\| < \left\|\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right\|$; it was also shown that $\hat{\boldsymbol{\beta}}_d$ has a better performance than $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ in the presence of multicollinearity.

In order to provide the explicit form of the MSE function of $\hat{\boldsymbol{\beta}}(k,d)$ use the following transformations:

Let $\boldsymbol{\alpha} = \mathbf{Q}'\boldsymbol{\beta}$, $\mathbf{Q}'\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{Q} = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{p+1})$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p+1} > 0$ such that the λ_j's are the eigenvalues of the matrix $\mathbf{X}'\mathbf{W}\mathbf{X}$ and $\mathbf{Q}$ is the matrix whose columns are the eigenvectors of the matrix $\mathbf{X}'\mathbf{W}\mathbf{X}$. ($\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{Q}'\boldsymbol{\Lambda}\mathbf{Q}$).

The MSE function of $\hat{\boldsymbol{\beta}}(k,d)$ is as follows:

$$\operatorname{MSE}\left(\hat{\boldsymbol{\beta}}(k,d)\right)=\sum_{j=1}^{p+1}\frac{\left(\lambda_j+kd\right)^2}{\lambda_j\left(\lambda_j+k\right)^2}+\sum_{j=1}^{p+1}\frac{k^2\left(1-d\right)^2\alpha_j^2}{\left(\lambda_j+k\right)^2} \tag{6}$$
$$=\mathrm{f}_1\left(k,d\right)+\mathrm{f}_2\left(k,d\right)$$

where $\mathrm{f}_1(k, d)$ is the variance function and $\mathrm{f}_2(k, d)$ is the squared bias. Thus choose suitable values for the parameters k and d in order to obtain a less MSE value than that of MLE.

Three theorems given in Huang (2012) are presented about the properties of the estimator $\hat{\boldsymbol{\beta}}(k,d)$:

Theorem 1. The asymptotic variance $\mathrm{f}_1(k, d)$ and the squared bias $\mathrm{f}_2(k, d)$ are two continuous functions of k and d; for fixed d^*, $0 < d^* < 1$, $\mathrm{f}_1(k, d^*)$ and $\mathrm{f}_2(k, d^*)$ are monotonically decreasing and increasing functions of k, respectively; for a fixed $k^* > 0$, $\mathrm{f}_1(k^*, d)$ and $\mathrm{f}_2(k^*, d)$ are monotonically increasing and decreasing functions of k, respectively (Huang, 2012).

Theorem 2. For a fixed d^*, $0 < d_* < 1$, there exists a $k > 0$ such that $\operatorname{MSE}\left(\hat{\boldsymbol{\beta}}(k,d)\right)<\operatorname{MSE}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right)$ (Huang, 2012).

Theorem 3. If $k > 0$ and $0 < d < 1$, then $\operatorname{MSEM}\left(\hat{\boldsymbol{\beta}}(k,d)\right)<\operatorname{MSEM}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right)$ if and only if $k(1-d)\boldsymbol{\alpha}'[k(1+d)\boldsymbol{\Lambda}^{-1}+2\mathbf{I}]^{-1}\boldsymbol{\alpha} < 1$, where $\operatorname{MSEM}\left(\tilde{\boldsymbol{\beta}}\right)=\mathrm{E}\left[\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\right]$, $\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}, \hat{\boldsymbol{\beta}}(k,d)$.

These theorems show the theoretical advantage of the estimator $\hat{\boldsymbol{\beta}}(k,d)$. In Huang (2012), the author designed a simulation study considering k = 0.1, 0.3, and 0.5 and d = 0.2, 0.4, and 0.6. However, it is anticipated that optimal value of d and some estimators of k will result in better performance.

Note that the values of the parameter d are restricted to the interval $0 < d < 1$ in the definition of $\hat{\boldsymbol{\beta}}(k,d)$ in order to obtain a sufficient condition satisfying $\operatorname{MSE}\left(\hat{\boldsymbol{\beta}}(k,d)\right)<\operatorname{MSE}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right)$. However, in the simulation given in Huang (2012), the author states that MSE of $\hat{\boldsymbol{\beta}}(k,d)$ can be smaller than MSE of $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ without satisfying this sufficient condition. Thus, we expand the restriction on d and conduct our simulation study such that the optimal parameter d_{opt} satisfies the following conditions: $-\infty < d_{\text{opt}} < \infty$, $d_{\text{opt}} \neq 0$, and $d_{\text{opt}} \neq 1$. In the following theorem, d_{opt} is presented.

Theorem 4. The optimal shrinkage parameter $-\infty < d_{\text{opt}} < \infty$ for minimizing $\text{MSE}\left(\hat{\boldsymbol{\beta}}(k,d)\right)$ $\forall k > 0$ is the following:

$$d_{\text{opt}} = \frac{\sum_{j=1}^{p+1} \frac{k\alpha_j^2 - 1}{\left(\lambda_j + k\right)^2}}{\sum_{j=1}^{p+1} \frac{k\left(1 + \lambda_j \alpha_j^2\right)}{\lambda_j \left(\lambda_j + k\right)^2}} \tag{7}$$

Proof: It is easy to find the optimal parameter d_{opt} by differentiating $\text{MSE}\left(\hat{\boldsymbol{\beta}}(k,d)\right)$ with respect to d and equating the derivative to zero. Solving the equation for d, we get the optimal parameter d_{opt}. ■

After choosing the optimal parameter d_{opt} for d, the parameter k must be selected. In literature, there are some estimators for the selection of k. The followings are the estimators of k that are used in the simulation study:

1. $k_1 = \dfrac{p+1}{\hat{\boldsymbol{\beta}}'_{\text{MLE}} \hat{\boldsymbol{\beta}}_{\text{MLE}}}$ (Schaefer et al., 1984)
2. $k_2 = \dfrac{p}{\hat{\boldsymbol{\beta}}'_{\text{MLE}} \hat{\boldsymbol{\beta}}_{\text{MLE}}}$ (Schaefer et al., 1984)
3. $k_3 = \dfrac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}$, where $\hat{\alpha}_{\max}$ is the maximum element of $\hat{\boldsymbol{\alpha}}_{\text{MLE}} = \mathbf{Q}' \hat{\boldsymbol{\beta}}_{\text{MLE}}$ and $\hat{\sigma}^2 = (\mathbf{y} - \hat{\mathbf{y}})' (\mathbf{y} - \hat{\mathbf{y}}) \big/ n - p - 1$ (Hoerl & Kennard, 1970)
4. $k_4 = \dfrac{1}{\hat{\alpha}_{\max}^2}$, which is a modified version of k_3 (Schaefer et al., 1984)
5. $k_5 = \dfrac{\hat{\sigma}^2}{\left(\prod_{j=1}^{p+1} \hat{\alpha}_j^2\right)^{\frac{1}{p+1}}}$, which is the geometric mean of $k_j = \hat{\sigma}^2 \big/ \hat{\alpha}_j^2$ (Kibria, 2003)

6. $k_6 = \text{median}\left(\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}\right)$ (Kibria, 2003)

7. $k_7 = \max\left(\frac{\lambda_j \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}\right)$ (Alkhamisi, Khalaf, & Shukur, 2006)

8. $k_8 = \max\left(\frac{1}{\sqrt{\hat{\sigma}^2 / \hat{\alpha}_j^2}}\right)$ (Muniz & Kibria, 2009)

Because the α_j^2's and σ^2 are not known in practice, the estimators $\hat{\alpha}_j^2$ and $\hat{\sigma}^2$ are used in the above formulae.

Monte Carlo Simulation

The effective factors are chosen to be the number of explanatory variables p, the sample size n, and the correlation among the explanatory variables ρ^2. MSE and mean absolute error (MAE) are used as the criterion of judgment.

The average MSE and MAE of the estimators $\hat{\boldsymbol{\beta}}_{\text{MLE}}$ and $\hat{\boldsymbol{\beta}}(k, d_{\text{opt}})$ for $k = k_1, \ldots, k_8$ are computed by using the following equations:

$$\text{AMSE}\left(\hat{\boldsymbol{\beta}}\right) = \sum_{r=1}^{3000} \frac{\left(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right)}{3000} \tag{8}$$

$$\text{MAE}\left(\hat{\boldsymbol{\beta}}_r\right) = \sum_{r=1}^{3000} \frac{\left|\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right|}{3000}\ , \tag{9}$$

where $\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}}_{\text{MLE}}, \hat{\boldsymbol{\beta}}(k, d_{\text{opt}})$ at the r^{th} step of the simulation.

Following Kibria (2003), in order to generate the explanatory variables, the following equation is used:

$$x_{ij} = \left(1-\rho^2\right)^{\frac{1}{2}} z_{ij} + \rho z_{ip} \tag{10}$$

where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p + 1$, and the z_{ij}'s are pseudo-random numbers following the standard normal distribution.

The dependent variable is obtained by using Be($\mathbf{P}$), where $P_i = {e^{x_i\boldsymbol{\beta}}} \big/ {1 + e^{x_i\boldsymbol{\beta}}}$ such that x_i is the i^{th} row of the design matrix $\mathbf{X}$. The parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_{p+1}$ are chosen due to Newhouse and Oman (1971) such that $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$, which is a commonly used restriction in many simulation studies in the field; for example see Alkhamisi and Shukur (2008), Asar, Karaibrahimoğlu, and Genç (2014), and Kibria (2003).

The following cases are considered: $\rho^2 = 0.90$, 0.95, and 0.99, $n = 50$, 100, and 200, and $p = 4$, 8, and 12. The simulation is repeated 3000 times for each set of (ρ^2, n, p). Thus, via this set up, it may be determined which of the estimators $k_1, \ldots, k_8$ has better performance when d_{opt} is used for different combinations of (ρ^2, n, p).

Results and Discussion

Results of the Monte Carlo simulation are compiled in Tables 1-6. The factors affecting the MSE's of the estimators in the simulation are the degree of correlation ρ, the sample size n, and the number of explanatory variables p.

Tables 1, 3, and 5 are the AMSE values of Liu-type estimators, MLE for different values of k, and optimal shrinkage parameter d_{opt} when $p = 4$, 8, and 12. According to these tables, when n and p are fixed, the increase in the correlation ρ causes an increase in the AMSE values of the estimators without exception. When $p = 12$, the increase in the correlation inflates the AMSE values drastically. The worst case is obtained when the sample size is low and the degree of correlation is high, namely, $n = 50$ and $\rho = 0.99$. When the sample size n is increased, fixing p and ρ, it has a positive effect on estimators; in other words, it can be seen that the AMSE values decreases for all of the estimators. Especially for MLE, there is a rapid decrease in the case of high correlation. When fixing n and ρ, if the number of explanatory variables is focused upon, it is observed that an increase in the value of p corresponds to an increase in the AMSE values for both MLE and the other estimators. When the number of explanatory variables is increased, one should also increase the sample size sufficiently in order to make stable estimations. The estimator having best performance among others is k_8 for all of the situations.

In Tables 2, 4 and 6, the MAE values are presented for the estimators for different values of k and optimal shrinkage parameter d_{opt} when $p = 4$, 8, and 12. Similar comments apply in the case of AMSE. However, the only difference is that the MAE values are significantly smaller than AMSE values for all of the cases.

Particularly for the case $p = 12$, the MAE values are much smaller than the AMSE values. Again, k_8 is the best option if MAE is used as a performance criterion.

Table 1. The AMSE values for different k with optimal d for $p = 4$

ρ	0.9			0.95			0.99		
n	50	100	200	50	100	200	50	100	200
k_1	1.1453	0.5309	0.4556	2.4285	0.9978	0.6572	10.5534	3.0258	1.5615
k_2	1.1650	0.5440	0.4647	2.4853	1.0291	0.6742	10.7168	3.0939	1.5964
k_3	1.3392	0.6530	0.5187	3.0932	1.2907	0.7937	12.8152	4.1027	2.0750
k_4	1.2258	0.5760	0.4791	2.6704	1.1247	0.7122	11.3301	3.3589	1.7338
k_5	1.2328	0.5734	0.4712	2.7752	1.1296	0.7079	11.3754	3.4391	1.7882
k_6	1.2192	0.6070	0.4969	2.5722	1.1291	0.7425	10.1046	3.0120	1.6530
k_7	1.2911	0.6498	0.5196	2.8291	1.2536	0.7898	10.5026	3.3305	1.8326
k_8	1.0532	0.5352	0.4600	2.1244	0.8674	0.6469	9.4613	2.6145	1.3765
MLE	3.1825	1.1528	0.8809	7.5693	2.6893	1.3873	31.3891	11.2257	5.3554

Table 2. The MAE values for different k with optimal d for $p = 4$

ρ	0.9			0.95			0.99		
n	50	100	200	50	100	200	50	100	200
k_1	0.9224	0.7068	0.6605	1.1921	0.9009	0.7785	2.1122	1.3928	1.1036
k_2	0.9319	0.7155	0.6674	1.2168	0.9187	0.7895	2.1498	1.4144	1.1182
k_3	1.0211	0.7872	0.7069	1.4623	1.0578	0.8626	2.6491	1.7357	1.3122
k_4	0.9613	0.7367	0.6779	1.2913	0.9696	0.8129	2.2942	1.5019	1.1735
k_5	0.9685	0.7344	0.6712	1.3379	0.9714	0.8085	2.3356	1.5334	1.1979
k_6	0.9688	0.7583	0.6912	1.2818	0.9809	0.8331	2.1056	1.4233	1.1560
k_7	1.0045	0.7861	0.7078	1.3853	1.0441	0.8613	2.2242	1.5346	1.2311
k_8	0.8891	0.7134	0.6655	1.0748	0.8401	0.7774	1.9088	1.2657	1.0335
MLE	1.6172	1.0400	0.9183	2.3961	1.5279	1.1320	4.5867	3.0125	2.1434

Table 3. The AMSE values for different k with optimal d for $p = 8$

ρ	0.9			0.95			0.99		
n	50	100	200	50	100	200	50	100	200
k_1	3.6513	1.3095	0.6274	9.2806	1.8112	0.9489	20.6034	9.796	2.5225
k_2	3.6891	1.3368	0.6433	9.3837	1.8477	0.9729	20.8967	9.9669	2.584
k_3	4.8402	2.0404	0.9534	12.8513	2.8887	1.5463	32.2306	15.7924	4.8666
k_4	4.0841	1.5948	0.7691	10.6056	2.2816	1.1993	24.3956	11.7169	3.3359
k_5	3.9290	1.5937	0.7289	10.0599	2.1088	1.1405	21.4002	11.2511	3.1836
k_6	3.7811	1.6056	0.8181	9.0515	2.0722	1.2374	18.5028	9.2001	2.788
k_7	4.1232	1.9067	0.9507	9.7368	2.5571	1.4772	18.6701	10.38	3.6014
k_8	3.2690	1.1391	0.6451	8.1544	1.5275	0.8871	17.7195	8.0577	1.9944
MLE	10.6802	4.1897	1.6451	30.2246	6.2970	2.7891	75.5528	35.5189	10.9108

Table 4. The MAE values for different *k* with optimal *d* for *p* = 8

ρ	0.9			0.95			0.99		
n	**50**	**100**	**200**	**50**	**100**	**200**	**50**	**100**	**200**
k_1	1.2863	1.0359	0.7590	2.0771	1.1701	0.9270	3.1881	2.3832	1.3794
k_2	1.2997	1.0486	0.7693	2.1008	1.1847	0.9396	3.2277	2.4142	1.3991
k_3	1.7004	1.3507	0.9518	2.8578	1.5745	1.2091	4.7419	3.4694	2.066
k_4	1.4414	1.1654	0.8462	2.3822	1.3516	1.0517	3.7236	2.759	1.6368
k_5	1.4218	1.1648	0.8202	2.3309	1.2945	1.0236	3.4021	2.7069	1.6007
k_6	1.4011	1.1833	0.8773	2.1406	1.3002	1.0766	3.0293	2.3767	1.5059
k_7	1.5140	1.3062	0.9517	2.3067	1.4789	1.1836	3.0591	2.6345	1.7617
k_8	1.2016	0.9769	0.7801	1.8728	1.0800	0.9094	2.8762	2.0945	1.2177
MLE	2.6659	1.9465	1.2458	4.5869	2.3624	1.6143	7.4523	5.4173	3.1146

Table 5. The AMSE values for different *k* with optimal *d* for *p* = 12

ρ	0.9			0.95			0.99		
n	**50**	**100**	**200**	**50**	**100**	**200**	**50**	**100**	**200**
k_1	139.2714	4.5505	0.8339	379.1795	18.6147	1.1519	693.7565	29.1359	5.4585
k_2	139.5613	4.5923	0.8505	379.7215	18.7365	1.1746	695.7493	29.472	5.5575
k_3	148.6546	6.4653	1.5089	400.7071	24.0658	2.3208	768.2535	48.479	11.1368
k_4	144.0903	5.4530	1.1198	389.2163	20.9824	1.6189	723.2426	35.3703	7.631
k_5	134.0508	4.9199	1.0217	367.5066	19.4268	1.5013	662.6167	32.3825	6.7584
k_6	131.8453	4.8675	1.1637	362.3602	17.9196	1.6392	642.9656	26.1222	5.7143
k_7	134.5658	5.8933	1.4827	364.9559	20.5335	2.1233	643.2291	27.9867	7.7012
k_8	129.5472	3.9218	0.8216	359.3362	16.0274	1.0576	639.7481	23.8502	4.1557
MLE	351.1918	11.6188	2.6647	935.6592	43.8108	4.1312	2048.4312	98.5518	22.1795

Table 6. The MAE values for different *k* with optimal *d* for *p* = 12

ρ	0.9			0.95			0.99		
n	**50**	**100**	**200**	**50**	**100**	**200**	**50**	**100**	**200**
k_1	8.4860	1.5200	0.8701	12.8861	2.8799	1.0276	17.5892	3.7843	1.9859
k_2	8.5038	1.5332	0.8794	12.9122	2.9003	1.0382	17.6425	3.8241	2.0081
k_3	9.1525	2.0939	1.1998	13.9870	3.8298	1.4903	19.9062	5.9098	3.1408
k_4	8.8189	1.7987	1.0191	13.4347	3.2974	1.2276	18.4166	4.5498	2.4684
k_5	8.3426	1.6721	0.9695	12.6869	3.1359	1.1809	17.4483	4.3274	2.3113
k_6	8.2283	1.6989	1.0465	12.4874	2.9680	1.2456	16.7388	3.6399	2.1214
k_7	8.4317	1.9781	1.1911	12.6229	3.3975	1.4279	16.7299	3.9188	2.5582
k_8	8.0607	1.3933	0.8786	12.2928	2.5904	0.9990	16.5748	3.2482	1.7039
MLE	13.9272	2.8855	1.5911	21.3777	5.3458	1.9804	33.1953	8.6407	4.4523

Application

An empirical application is demonstrated by using a data set taken from the web site of Statistics Sweden[1]. There are 290 municipalities in Sweden; eighty-three municipalities are considered. Those considered are the urban regions defined as the municipalities belonging to the Functional analysis regions Stockholm, Göteborg and Malmö, corresponding to the year 2012. The explanatory variables are defined as follows: X_1 is the population, X_2 is the number of unemployed people, X_3 is the number of newly constructed buildings, and X_4 is the number of bankrupt firms. A binary logistic regression model is set by using the dependent variable defined as follows: If there is an increase in the population of a municipality it is coded as 1; otherwise it is coded as 0.

It is observed from Table 7 that the bivariate correlations among the regressors are high (all greater than 0.91) and the condition number of the data given by $\kappa = \sqrt{\lambda_{\max} / \lambda_{\min}} = 38.3274$ shows that there is severe multicollinearity problem with this data.

For different values of k defined above and for the optimal shrinkage parameter d_{opt}, the MSEs of the estimators are given in Table 8. It can be seen from that table that k_8 has the best performance in the sense of MSE reduction.

Also, k_5 and k_6 perform quite well. If the coefficients, standard errors of the estimators, and the corresponding t-values given in Table 9 are considered, it is seen that k_5 and k_6 have very low standard errors when compared to the other estimators. Moreover, t-values corresponding to k_5 and k_6 are larger than the others in absolute value which further shows the superiority of k_5 and k_6. Thus, k_5 and k_6 seem to be more practical for this data set. Finally, a graph of the MSE function versus k is provided in Figure 1. According to Figure 1, the MSE of $\hat{\boldsymbol{\beta}}\left(k, d_{\text{opt}}\right)$ has a decreasing tendency for the increasing values of the parameter k.

Table 7. The correlation matrix of the data used in the application

	X_1	X_2	X_3	X_4
X_1	1.0000	0.9937	0.9707	0.9514
X_2	0.9937	1.0000	0.9527	0.9222
X_3	0.9707	0.9527	1.0000	0.9765
X_4	0.9514	0.9222	0.9765	1.0000

Table 8. The MSEs of the estimators used in the application

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	MLE
MSE	766.7641	776.6589	850.1244	813.5330	724.3047	693.2426	732.2801	687.6025	1894.307

Table 9. Coefficients, standard errors, and t-values of the data

	Coefficients								
	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	MLE
β_1	9.5174	9.5364	9.7763	9.6347	9.4529	9.3549	9.4654	9.1923	25.3187
β_2	-5.1346	-5.1261	-5.3010	-5.1732	-5.3543	-5.9703	-5.2813	-6.3168	-17.4101
β_3	2.3166	2.4417	3.3105	2.8941	1.7992	1.5479	1.8892	1.4059	3.8671
β_4	-5.7236	-5.8749	-6.8082	-6.3768	-4.9382	-4.0628	-5.1086	-3.9790	-10.9671
	Standard errors								
	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	MLE
β_1	1.2685	1.2749	1.3382	1.3035	1.2496	1.2513	1.2519	1.2550	3.4576
β_2	1.0505	1.0579	1.1134	1.0852	1.0167	0.9884	1.0235	0.9840	2.7109
β_3	0.8606	0.9019	1.1718	1.0450	0.6694	0.5164	0.7069	0.4866	1.3404
β_4	0.7752	0.8023	0.9745	0.8942	0.6404	0.5067	0.6686	0.4764	1.3124
	t-values								
	k1	k2	k3	k4	k5	k6	k7	k8	MLE
β_1	7.5029	7.4804	7.3054	7.3911	7.5650	7.4761	7.5609	7.3244	7.3226
β_2	-4.8877	-4.8458	-4.7611	-4.767	-5.2665	-6.0405	-5.1599	-6.4195	-6.4222
β_3	2.6919	2.7073	2.8252	2.7694	2.6878	2.9977	2.6726	2.8891	2.8850
β_4	-7.3835	-7.3224	-6.9862	-7.1315	-7.7113	-8.0179	-7.6402	-8.3516	-8.3567

Summary and Conclusion

The benefits of Liu-type estimators in logistic regression were shown in the case of multicollinearity. The optimal shrinkage parameter d_{opt} used in the Liu-type estimator was defined in Huang (2012). The Monte Carlo experiment was used to evaluate the early proposed ridge regression parameters in logistic regression. Results show that the estimators chosen from the literature outperform MLE for all of the cases taken into consideration when the optimal shrinkage parameter d_{opt} is used. AMSE and MAE values of MLE become inflated when the correlation increases and the sample size decreases. Thus, researchers are advised to use Liu-type logistic estimators with the optimal shrinkage parameter in place of MLE in the presence of multicollinearity. The best performance from the estimator will be obtained if k_8 is used.

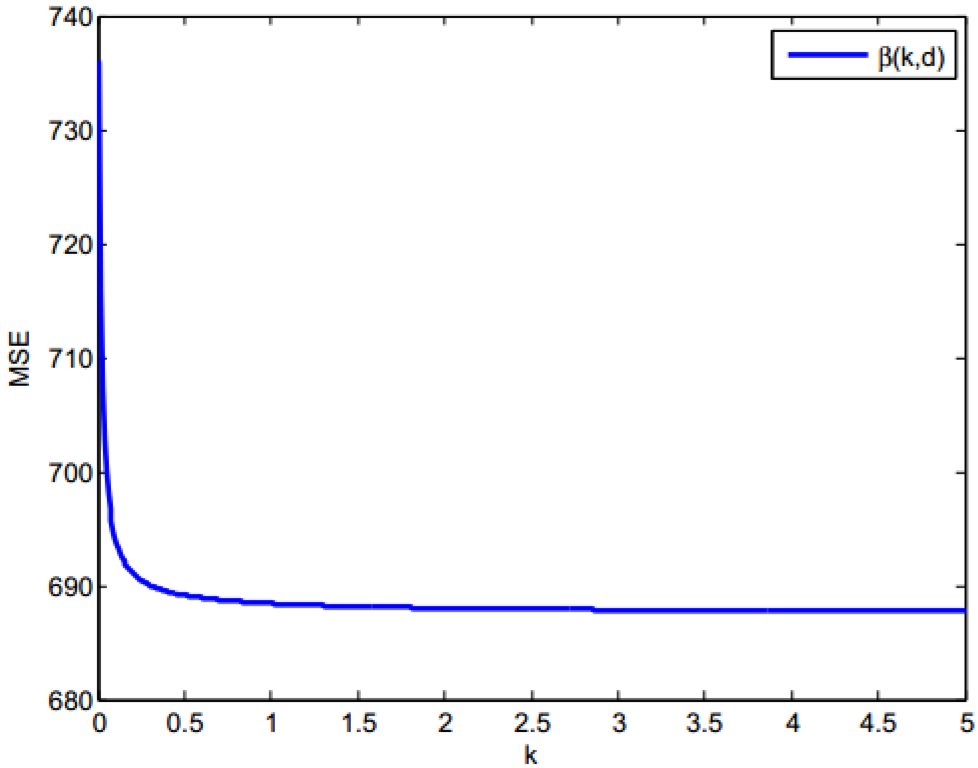

Figure 1. The MSE of $\hat{\boldsymbol{\beta}}(k,d)$ function versus *k*

Footnotes

1. The data used in this article may be accessed through the following website: http://www.scb.se/sv_/Hitta-statistik/

References

Alkhamisi, M., Khalaf, G., & Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics-Theory and Methods, 35*(11), 2005-2020. doi: 10.1080/03610920600762905

Alkhamisi, M., & Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics-Theory and Methods, 37*(4), 544-564. doi: 10.1080/03610920701469152

Asar, Y., Karaibrahimoğlu, A., & Genç, A. (2014). Modified ridge regression parameters: A comparative Monte Carlo study. *Hacettepe Journal of Mathematics and Statistics, 43*(5), 827-841. Retrieved from http://www.hjms.hacettepe.edu.tr/uploads/7742744a-b036-456d-8367-1a3eab73e172.pdf

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67. doi: 10.1080/00401706.1970.10488634

Huang, J. (2012). A simulation research on a biased estimator in logistic regression model. In Z. Li, X. Li, Y. Liu, & Z. Cai (Eds.), *Computational Intelligence and Intelligent Systems* (389-395). Berlin, Germany: Springer. doi: 10.1007/978-3-642-34289-9_43

Inan, D., & Erdogan, B. E. (2013). Liu-type logistic estimator. *Communications in Statistics-Simulation and Computation, 42*(7), 1578-1586. doi: 10.1080/03610918.2012.667480

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation, 32*(2), 419-435. doi: 10.1081/SAC-120017499

Kibria, B. M. G., Månsson, K., & Shukur, G. (2012). Performance of some logistic ridge regression estimators. Computational Economics, 40(4), 401-414. doi: 10.1007/s10614-011-9275-x

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods, 22*(2), 393-402. doi: 10.1080/03610929308831027

Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods*, *32*(5), 1009-1020. doi: 10.1081/STA-120019959

Månsson, K., Kibria, B. M. G., & Shukur, G. (2012). On Liu estimators for the logit regression model. *Economic Modelling*, *29*(4), 1483-1488. doi: 10.1016/j.econmod.2011.11.015

Månsson, K., & Shukur, G. (2011). On ridge parameters in logistic regression. *Communications in Statistics-Theory and Methods, 40*(18), 3366-3381. doi: 10.1080/03610926.2010.500111

Muniz, G., & Kibria, B. M. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics-Simulation and Computation, 38*(3), 621-630. doi: 10.1080/03610910802592838

Newhouse, J. P., & Oman, S. D. (1971). An evaluation of ridge estimators: A report prepared for the United States Air Force Project Rand (P-716-PR). Retrieved from http://www.rand.org/content/dam/rand/pubs/reports/2007/R716.pdf

Schaefer, R. L., Roi, L. D., & Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics-Theory and Methods*, *13*(1), 99-113. doi: 10.1080/03610928408828664